

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE (DD-MM-YYYY) 22-08-2014		2. REPORT TYPE Final		3. DATES COVERED (From - To) 7 July 2013 – 22 July 2014	
4. TITLE AND SUBTITLE  Evaluation of Tests of Perceptual Speed/Accuracy and Spatial Ability for Use in Military Occupational Classification				5a. CONTRACT NUMBER In-House	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62202F	
6. AUTHOR(S)  Janet D. Held <sup>1</sup> , Thomas R. Carretta <sup>2</sup> , and Michael G. Rumsey <sup>3</sup>				5d. PROJECT NUMBER 5329	
				5e. TASK NUMBER 09	
				5f. WORK UNIT NUMBER 53290902	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) See back of this form.				8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711 Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Supervisory Control and Cognition Branch Wright-Patterson AFB OH 45433				10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHCI	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distriubution A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES 88 ABW Cleared 08/22/2014; 88ABW-2014-3943.					
14. ABSTRACT With the exception of Assembling Objects (AO), a spatial ability test used only by the Navy in enlisted occupational classification, the Armed Services Vocational Aptitude Battery (ASVAB) is academic and knowledge-based, somewhat limiting its utility for occupational classification. This article presents the case for integrating the AO test into military classification composites and for expanding the breadth of ASVAB content by including a former ASVAB speed/accuracy test, Coding Speed (CS). Empirical evidence is presented that shows AO and CS (increment the validity of the ASVAB in predicting training grades for a broad range of occupations, (b) reduce adverse impact defined as test score barriers for women and minorities, and (c) improve classification in terms of matching recruits to occupations. Some cognitive theory is presented to support AO and CS, as well as nonverbal reasoning and working memory tests for inclusion in or adjuncts to the ASVAB.					
15. SUBJECT TERMS ASVAB, incremental validity, adverse impact, classification effectiveness, Coding Speed, Assembling Objects					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  SAR	18. NUMBER OF PAGES  24	19a. NAME OF RESPONSIBLE PERSON Antonio Ayala
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

## 7. PERFORMING ORGANIZATIONS NAMES AND ADDRESSES:

Janet D. Held<sup>1</sup>

Bureau of Naval Personnel

Navy Personnel Research, Studies, and Technology

Millington, TN

Thomas R. Carretta<sup>2</sup>

Air Force Research Laboratory

Wright-Patterson AFB, OH

Michael G. Rumsey<sup>3</sup>

Annandale, VA

## Evaluation of Tests of Perceptual Speed/Accuracy and Spatial Ability for Use in Military Occupational Classification

Janet D. Held

Navy Personnel, Research, Studies, and Technology  
Millington, Tennessee

Thomas R. Carretta

Air Force Research Laboratory,  
Wright-Patterson Air Force Base, Ohio

Michael G. Rumsey

Annandale, Virginia

With the exception of Assembling Objects (AO), a spatial ability test used only by the Navy in enlisted occupational classification, the Armed Services Vocational Aptitude Battery (ASVAB) is academic and knowledge-based, somewhat limiting its utility for occupational classification. This article presents the case for integrating the AO test into military classification composites and for expanding the breadth of ASVAB content by including a former ASVAB speed/accuracy test, Coding Speed (CS). Empirical evidence is presented that shows AO and CS (a) increment the validity of the ASVAB in predicting training grades for a broad array of occupations, (b) reduce adverse impact defined as test score barriers for women and minorities, and (c) improve classification in terms of matching recruits to occupations. Some cognitive theory is presented to support AO and CS, as well as nonverbal reasoning and working memory tests for inclusion in or adjuncts to the ASVAB.

**Keywords:** ASVAB, incremental validity, adverse impact, classification effectiveness, coding speed, assembling objects

The Armed Services Vocational Aptitude Battery (ASVAB) is used by all of the U.S. military services for enlistment qualification and to classify enlistees into military occupations. Because some military jobs change over time, joint-service collaborations have occurred, researching how to augment the breadth of the domain/constructs measured by the ASVAB. The current battery predominately measures academic achievement (math, verbal, science) and technical knowledge (mechanical,

electronics, auto/shop). Although the ASVAB contains tests designed to measure the aptitude domain related to training performance in military jobs, much of its content is also linked to job knowledge and job performance constructs. Strengthening the relationship between the aptitude/ability/learning capabilities measured by the ASVAB with military performance improves the ability to more accurately assign individuals to occupations for which they are likely to succeed, therefore lowering military costs and the personal costs associated with failure.

All of the services have conducted personnel selection and classification research over the years, with a major objective of expanding the ASVAB. The most comprehensive effort was the Army's Project A, which expanded not only the predictor domain but also the military performance domain, or criterion space, upon which the predictors would be validated (Buscigilo, Palmer, King, & Walker, 1994; Campbell & Zook, 1992; Russell & Peterson, 2001). Another large research effort was a joint-

---

Janet D. Held, Navy Personnel Research, Studies, and Technology, Bureau of Naval Personnel, Millington, Tennessee; Thomas R. Carretta, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio; Michael G. Rumsey, Annandale, Virginia.

The opinions expressed are those of the authors and not necessarily those of the U.S. Government, Department of Defense, the U.S. Navy, or the U.S. Air Force.

Correspondence concerning this article should be addressed to Janet D. Held, Navy Personnel Research, Studies, and Technology, Bureau of Naval Personnel, 5720 Integrity Drive, Millington, TN 38055-1000. E-mail: [janet.held@navy.mil](mailto:janet.held@navy.mil)

service project that capitalized on the technology that launched the computer adaptive version of the ASVAB, the CAT-ASVAB. This project was known as the Enhanced Computer-Administered Test (ECAT) battery (Alderton, Wolfe, & Larson, 1997). The ECAT project was “. . . driven by cognitive theories of aptitude, working memory, and mental imagery” (Alderton et al., 1997, p. 7), specifically, Carroll’s (1993) theory of cognitive abilities. At the culmination of the ECAT project, only one test was chosen for addition to the ASVAB: the spatial ability test, Assembling Objects (AO). The major reasons at the time for selecting the AO test for the ASVAB were a small but meaningful degree of incremental validity for some of the studied occupations and the demonstration of reduced adverse impact, but also that the test could be administered in both paper-and-pencil and computer formats, whereas other ECATs could not. For a full discussion of the ECAT battery, see the special 1997 *Military Psychology* issue (Volume 9, Number 1) dedicated to the ECAT.

With regard to ASVAB and ECAT construct overlap and unique ECAT construct measurement, the Navy conducted several factor analyses (Alderton et al., 1997) that varied in extraction method, rotation method, number of factors extracted, and initial communality estimates. The most representative structure came from a hierarchical factor solution favoring Carroll’s (1993) structure of a general ability factor and orthogonal (unrelated) specific abilities. Factor analyzed as separate batteries, the ASVAB showed an overarching general ability factor with four clear lower-level factors of Technical Knowledge, Verbal Ability, Clerical Speed (which contained the Numerical Operations [NO] and Coding Speed [CS] tests), and Mathematics Ability. In contrast, the ECAT, which also had an overarching general ability factor, showed different lower-level factors than were observed in the ASVAB: Spatial (which contained AO, among other tests measuring spatial ability), Psychomotor, and Working Memory.

More recently, an external ASVAB review panel with expertise in personnel selection, job classification, psychometrics, and cognitive psychology met to consider the current ASVAB’s content and the testing research conducted by the military personnel research laboratories (Drasgow, Embretson, Kyllonen, &

Schmitt, 2006). As part of the panel’s evaluation, Drasgow and colleagues also examined ASVAB content in light of Carroll’s (1993) stratum theory of the structure of intellect and conducted confirmatory factor analyses of the ASVAB tests based on the Spearman-Holzinger bifactor model (Holzinger & Harman, 1941). Drasgow and colleagues found, as have others (e.g., Ree & Carretta, 1994), a strong general factor for the ASVAB dominated by the verbal and math tests, which they interpreted as crystallized intelligence (Gc). Crystallized intelligence loads strongly on language skills (e.g., vocabulary) and education (general and specific knowledge) and therefore reflects intellectual achievement that, in turn, depends somewhat on access to quality education or specialized knowledge. Gc may also be linked to socioeconomic status, interests, and/or opportunity. In contrast, the ECAT is considered by psychologists to measure fluid intelligence (Gf). Gf can be described as the ability to think logically and solve problems in novel situations independent of knowledge acquired through education, learning, or experience. Given the increasingly diverse youth population and the emphasis of several emerging military occupations on the ability to think logically and solve problems (e.g., cyber occupations), it seems appropriate for the ASVAB to contain more measures of fluid intelligence than just AO.

CS is a former ASVAB test that is currently administered as a special classification test to Navy applicants. Although CS is under the umbrella of Gf, it may be viewed more as a process-based or processing-perceptual speed test, where performance depends on the speed and accuracy with which individuals perform simple information processing tasks (Ackerman & Cianciolo, 2000). By process-based, we mean that the test content is uncomplicated and incidental to the ability being measured. Process-based measures like CS that do not rely on learned content have contributed to military personnel selection batteries since World War I. Dockeray and Isaacs (1921) reported that both Italy and France included measures of reaction time (RT) in their pilot selection batteries, a slightly different construct than CS, but nevertheless measuring speed. Thurstone’s work on the identification of primary mental abilities (Thurstone, 1938; Thurstone & Thurstone, 1941) provided further support for the impor-

tance of process-based measures and Gf in their identification of perceptual speed, memory, and space factors.

In support of augmenting the ASVAB content, many theorists have proposed that including and differentially weighting tests that measure specific abilities that are important for occupational areas should result in better prediction of occupational performance than merely depending on measures of general cognitive ability. This hypothesis is referred to as specific aptitude theory or differential aptitude theory (Hull, 1928; Thurstone, 1938). The influence of differential aptitude theory is reflected in the development of taxonomies of human abilities (e.g., Fleishman, Quaintance, & Broadling, 1994) and military multiple-aptitude test batteries that are somewhat an outgrowth of these taxonomies. For example, tests of spatial ability and processing speed that do not rely on learned content have been a mainstay of multiple aptitude aircrew test batteries (Carretta & Ree, 2003) such as the Air Force Officer Qualifying Test (Dragow, Nye, Carretta, & Ree, 2010) and other aircrew aptitude batteries (Carretta & Ree, 2003) for many years, as well as periodically appearing on the ASVAB and service-specific batteries (Rumsey, 2012). Further, with regard to speeded tests, Alf and Gordon (1957) demonstrated the broader application of

so-called simple clerical tests for military occupations when they found a Navy clerical composite had higher validity for predicting Navy frogmen (early designation for Navy SEALs [Sea, Air, and Land]) fleet performance ( $r = .40$ ) than did knowledge-based tests.

The influence of differential aptitude theory is still pervasive and has been adopted by the Army in the development of differential assignment theory (DAT; Johnson & Zeidner, 1995; Zeidner & Johnson, 1991, 1994). DAT is a multifaceted theoretical framework firmly grounded in classification principles that considers both predictive validity (stressed in general mental ability theories) and differential validity (specific ability measures contributing incremental validity over general mental ability). The application of DAT is intended to improve the process of optimally matching people to jobs and has been incorporated into the Army's enlisted personnel classification algorithm (Johnson & Zeidner, 1995; McWhite & Greenston, 1998). DAT is discussed later in the paper in the context of classification efficiency. Table 1 provides brief descriptions of the ASVAB that includes AO, but also the CS special classification test currently delivered only on the CAT-ASVAB platform to Navy applicants.

Table 1  
*Description of the ASVAB and Coding Speed Tests*

Test name and abbreviation	Test description
General Science (GS)	Knowledge of physical and biological sciences
Arithmetic Reasoning (AR)	Ability to solve arithmetic word problems
Word Knowledge (WK) <sup>a</sup>	Ability to select the correct meaning of words presented in context and correct synonyms
Paragraph Comprehension (PC) <sup>a</sup>	Ability to obtain information from written passages
Mathematics Knowledge (MK)	Knowledge of high school mathematics principles
Electronics Information (EI)	Knowledge of electricity and electronics
Auto and Shop Information (AS)	Knowledge of automobile and shop technologies tools and practices
Mechanical Comprehension (MC)	Knowledge of mechanical and physical principles
Assembling Objects (AO) <sup>b</sup>	Ability to determine correct spatial forms from their separate parts and connection points
Coding Speed (CS) <sup>b</sup>	Ability to quickly identify correct word/number pairings from a key with many options

*Note.* ASVAB = Armed Services Vocational Aptitude Battery.

<sup>a</sup> WK and PC are combined to form the Verbal (VE) composite that is a component of the AFQT and several Navy ASVAB classification composites. <sup>b</sup> Not all recruits enter the Navy with AO and CS test scores. CS is only given by computer at the MEPS at the end of the computer-administered CAT-ASVAB. AO, also given on the CAT-ASVAB, is not given to high school students taking the paper and pencil version of the ASVAB under the Career Exploration Program, but is given in paper-and-pencil ASVAB forms in the Enlisted Testing Program.

As explained earlier, the ASVAB has undergone several content changes since its implementation in the 1970s, and there is full support by the Department of Defense (DoD)/services for further change. Drasgow et al. (2006) had many recommendations regarding content of the battery. They recommended a review of the ASVAB content, a revisit of the ECAT results, and consideration of new content to include measures of noncognitive characteristics, a technical knowledge test of information/communications technology literacy, and enhanced measurement of nonverbal reasoning. The rationale for the inclusion of nonverbal reasoning tests includes expanding the breadth of the measurement of general mental ability, improving classification effectiveness, reducing adverse impact, and improving the assessment of cognitive ability and trainability in applicants challenged in English skills (e.g., non-native English speakers; Drasgow et al., 2006). The AO spatial ability test, depending on the type of factor analysis, can be considered somewhat of a nonverbal reasoning test and is the only such test included in the current ASVAB. The DoD is now preparing to evaluate several nonverbal reasoning tests including a working memory test. The CS test, a former ASVAB test, while not considered a nonverbal reasoning test, has its own merits and has been revamped to address some issues discussed in this paper.

### Purpose

The purpose of this paper is to provide the history of the CS and AO tests and the supporting theoretical and empirical evidence for their use in military occupational classification. Studies are reviewed that focus on (a) incremental validity when used in combination with the academic/technical knowledge-based ASVAB tests for predicting military training performance criteria, (b) reducing subgroup differences (adverse impact) for women and racial/ethnic minority groups, and (c) improved classification in terms of matching recruits to occupations. Although the analyses reported here are not exhaustive, they provide evidence of the utility of CS and AO and insights regarding the likely benefits of measures of reduced verbal content or process-based tests in supplementing the ASVAB verbal, math, and technical knowledge tests.

## History and Use of CS and AO in U.S. Military Personnel Selection and Classification

The following sections describe the history of the Coding Speed (CS) and Assembling Objects (AO) tests and the theoretical and empirical evidence supporting use in military occupational classification.

### Coding Speed

The Army developed the earliest-known U.S. military test of "coding speed" used operationally for military occupation classification (Helme, Graham, & Anderson, 1962). Helme and colleagues described the Army Clerical Speed Test, which closely resembles the former ASVAB CS test. The Navy subsequently modified the Clerical Speed Test and adopted it as part of their Basic Test Battery. In 1976, the first joint-service ASVAB forms (Forms 6 and 7) were introduced for enlistment qualification and classification and they contained a different clerical speed test, Attention to Detail (AD). The AD test was subsequently considered suboptimal in predictive validity and classification utility, so in October 1980, AD was replaced by CS in ASVAB Forms 8, 9, and 10.

From 1980 to 2002, the ASVAB contained two speeded tests, NO and CS, with CS used most widely in classifying military recruits to clerical occupations (Weltin & Popelka, 1983), but with the Army and Navy using the tests for a variety of occupations. Both tests were eliminated from the battery in 2002 because of problems associated with speeded tests; mainly, examinees' scores were sensitive to changes in test format and item response input modes. For example, in their paper-and-pencil format, NO and CS scores were impacted when the answer sheet with round bubbles for marking responses was replaced with one that had narrow vertically placed rectangles (Bloxom, Thomasson, Wise, & Branch, 1993; Ree & Wegner, 1990). The answer sheet with rectangles took less time to input responses than the answer sheet with circles because only one up-and-down stroke of the pencil was required to fill in the rectangles compared to the longer time it took to more carefully exercise a circular motion to fill in the circles. Given that the NO and CS tests were scored as number of items correct under a time



limit, examinees with the rectangle answer sheets had on average more correct responses than examinees with the bubble answer sheets.

NO and CS score impact issues again became a concern when the ASVAB became computer adaptive (CAT-ASVAB). To study potential score impacts on NO and CS, a CAT-ASVAB computer hardware effects study was designed with conditions that varied computer features (CPU, monitor size, response input device, color scheme, and portability; Segall, 1997). The study showed that NO was sensitive to both response input device (e.g., keypad vs. the existing template covered keyboard specially designed for CAT-ASVAB) and computer portability (e.g., subnotebook vs. desktop PC). The CS test was only sensitive to portability (with acknowledgment that the exact features that caused the score differences would be hard to determine), but it appeared that only the speed component was affected, not the accuracy component (Segall, 1997, p. 226). It should be noted that no statistically significant answer sheet effects or computer hardware effects were observed for the ASVAB power tests (Bloxom et al., 1993) and the ASVAB tests have since been considered robust to platform changes, including Internet delivery.

The Defense Manpower Data Center (DMDC) and the Navy paid particular attention to CS during the speeded tests' evaluation because of some of the documented benefits for enhancing the military classification systems discussed in this paper. One of DMDC's supporting efforts for both NO and CS was the development of a more robust "rate score" to replace the simple number of items answered correctly score (Segall, Moreno, Bloxom, & Hetter, 1997). The rate score is essentially the average per minute number of items correct corrected for guessing (e.g., fast random responding), factoring in an adequate screen display time. As explained by Segall et al. (1997, pp. 137–138), the rate score is more suitable for speeded tests where changes in aspects of the test and delivery platforms require consideration of the test time limit. A recent concern for CS was the replacement of the specially configured CAT-ASVAB keyboard with a mouse for response input. Mouse input was expected to produce faster responses because examinees would not need to look down at the keyboard for the correct key (A, B, C, D, or E) to press;

that is, the item choices for mouse response choices are displayed low on the computer screen to merely be clicked. At the time of this writing, DMDC had completed their study of CS score differences between response input modes and found no differences and thus no need for a special CS score equating (DMDC briefing given to Navy on October 30, 2013).

Aside from the CS rate score change, other improvements to the test have occurred over time to make CS a more robust test. One of the improvements, made by DMDC, involved simplifying the test's instructions, because there was evidence that some of the CS score variance was due to individual differences in the ability to understand them. Also during this CS review time, contract support focused on the computerized version of CS that included formatting changes (to more closely resemble the paper-and-pencil version) and increased opportunity to review the revised instructions and engage in the practice items (Abrahams et al., 1996). In addition, Abrahams and Alf (2001) compared several well-known perceptual speed tests that supported the construct validity of the CS test and found further support from the early work of Ghiselli (1966), which showed measures of perceptual speed were useful for predicting both training and job performance.

With all of the attention given to CS and the Navy's empirical evidence supporting the test, the Navy was able to retain CS as a special classification test administered seamlessly at the end of the CAT-ASVAB. In 2004, DMDC scaled the four computerized CS forms to the newest ASVAB normative population score scale (Segall, 2004). From that time to the present, CS has not shown indications of compromise or score drift even though the original paper-and-pencil items (four forms) were retained for the computerized version. CS is now administered to all Navy applicants testing on the CAT-ASVAB at all of the Military Entrance Processing Stations (MEPS), where the computer hardware features are not widely disparate. CS, however, is not administered at any of the Military Entrance Test (MET) sites. MET sites are generally more remotely located than MEPS, are lower volume, and do not administer special tests. In the last few years, the addition of computers with Internet connectivity has converted about 50% of the MET sites to Web-based administration of the CAT-ASVAB. The

CS test could be administered at these now-Web-based MET sites; however, DMDC would have to conduct another CS study to determine if Internet delivery of the items (that sometimes lags) impacts CS scores and if so, how to control for the effect. At this point, because not all Navy applicants are administered the CS or AO tests, all ratings that have as their operational classification composite one that includes either test must also have an alternative ASVAB composite that does not contain these tests.

### Assembling Objects

The Army also developed the AO test, but at a different time than CS. AO was developed during the Army's Project A (Buscigilo et al., 1994; Campbell & Zook, 1992; Russell & Peterson, 2001). One of the first steps in Project A was the identification of abilities and characteristics important to Army occupations that were not measured by the ASVAB. Spatial ability was identified as a key area. Several spatial constructs were identified; 10 spatial tests were developed, six of which survived field testing and were included in validation studies (Russell et al., 2001).

Factor analyses of the Project A spatial tests indicated the presence of a general spatial factor and that reasoning and assembly type items were the best measures of this factor. Additional analyses revealed that there were small or no gender differences for the spatial tests, AO and Figural Reasoning (FR; Peterson et al., 1990). Further, in a study of the effects of practice and coaching on test performance, only small-to-moderate mean score improvements were observed for AO and FR (Buscigilo & Palmer, 1996). Both AO and FR were included in the DoD's ECAT (Alderton et al., 1997) project. Analyses of the ECAT data showed that AO could increment the validity of the ASVAB for predicting job performance and improve classification of personnel into some military occupations (Sager, Peterson, Oppler, Rosse, & Walker, 1997; Wolfe, Alderton, Larson, Blom, & Wise, 1997). The AO test subsequently became an ASVAB test in 2002 when NO and CS were eliminated.

It should be noted that on a theoretical basis, the Navy, in their use of AO and CS, has not combined the two tests in the same classification composite. One reason is that the tests are

considered measurements of separate constructs linked to different occupations. AO measures the ability to visually construct spatial forms from the forms' parts and also to identify connection points of form parts. On the face of it, these types of test items map well to tasks performed in mechanical occupations (Held, Fedak, & Johns, 2004). CS, on the other hand, requires quick and accurate thinking, which applies to many operations types of occupations in addition to clerical (e.g., Navy SEALs). The AO and CS tests, in more comprehensive analyses, will be evaluated in combination in the future across a wide variety of military occupations.

The second reason for not combining AO and CS in the same classification composite is logistical in that not all Navy applicants are administered both tests. For example, Navy applicants testing on the paper-and-pencil version of the ASVAB receive AO but do not receive CS. Further, those who take the ASVAB in the high school testing program (Career Exploration Program, currently administering ASVAB in paper-and-pencil) do not receive either AO or CS.

The third reason for not combining AO and CS in the same composite is that initial validity analyses with data available for both tests were not supportive. For example, in the ECAT study, Held and Wolfe (1997) added the "best" ASVAB test to operational ASVAB classification composites (two to four tests in the composites) and compared the incremental validity with that provided by the best ECAT test. The ECAT incremental validity results showed AO did not add to the ASVAB operational composite for the six occupations that used CS in their ASVAB composite (Held & Wolfe, 1997, p. 81).

The Army has conducted extensive analyses on the AO test and found that it has potential to be included in their classification systems. Further, it has been suggested that AO could be used in a revised version of the ASVAB Armed Forces Qualification Test (AFQT<sup>1</sup>), which is

<sup>1</sup> AFQT scores are calculated from a linear combination of the ASVAB verbal (PC and WK) and math (AR and MK) standardized scores and are reported as percentiles.  $AFQT_s = S_{AR} + S_{MK} + 2S_{VE}$ , where VE (Verbal) is a weighted composite of the PC and WK tests (Segall, 1997). In addition to the AFQT, the services screen military applicants on education, mental, moral and physical factors.



used for military enlistment qualification (Anderson et al., 2011). We expect that the addition of AO to the AFQT is not likely but that the Army and the other services will find the AO test useful in occupational classification.

### Criteria for Evaluating the Utility of AO and CS

There are well-established professional guidelines regarding the development and use of tests for personnel measurement and selection (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Society for Industrial and Organizational Psychology, 2003). As stated earlier, in this paper, we focus on three test evaluation factors of particular importance to the U.S. military: (a) incrementing validity when used in combination with the ASVAB for predicting important performance criteria, (b) reducing subgroup differences (adverse impact) for women and at least some racial/ethnic minority groups, and (c) improved classification in terms of matching recruits to occupations.

When proposing new content be added to the ASVAB, Drasgow et al. (2006, p. 25) emphasized the potential benefits of reducing adverse impact/expanding the applicant pool and improving classification efficiency (CE). Drasgow et al. referred to CE in the overarching context of classification theory and the Army's DAT; that is, tapping into relevant aptitudes/abilities/skills that individuals do not have to the same degree and that apply more strongly to different occupational groups. Although Drasgow et al. downplayed the importance of incremental validity for new tests (e.g., measures could have the same predictive validity as currently observed for the ASVAB but may benefit the military in other ways), we provide evidence that incremental validity is achievable. In practice, however, we recognize that new test content may produce mixed results. For example, the addition of a psychomotor test to the ASVAB might provide incremental validity but increase adverse impact for women and also increase administration costs due to the requirement for specialized input devices, which was found to be the case in the evaluation of the ECAT psychomotor tests.

The decision to supplement the ASVAB with new content must include weighing positive and negative impacts on several factors, not just one or two. For example, as pertains to the focus of this paper, a new measure should demonstrate predictive validity for more than one occupation or the measure is not cost-effective, at least for broad applicant administration. Also, it would be desirable for the test to predict more than one performance measure, not just training grades (e.g., work samples, supervisor and peer ratings, and attrition/retention). The Navy, however, stresses prediction of performance in training as the most relevant criterion because so many of their ratings are technically complex and failure costs at this point are high. The Army has taken a more comprehensive approach and has led the way in the measurement of posttraining performance measurement, predominantly in Project A, but also in their more recent evaluation of noncognitive measures that map better to job performance than to training performance.

### Incremental Validity to the ASVAB

**Coding Speed.** Table 2 summarizes validity coefficients for predicting final school grades in training for several Navy ratings for ASVAB composites with and without CS. (Navy ratings are enlisted occupations similar to Army and Marine Corps military occupational specialties and Air Force specialties.) The validities in Table 2 were corrected for range restriction using the multivariate method (Lawley, 1943) as applied in the military ASVAB context (Held & Foley, 1994) but not criterion unreliability. Table 2 does not include two Navy composites that contained NO because the composites did not show incremental validity to the evaluation (baseline) composite,<sup>2</sup> VE + MK. During the DoD evaluation of the speeded tests, it was suggested that the baseline composite (VE + MK) was an adequate replacement for service clerical composites that contained either NO or CS.

The validity results in Table 2 were presented to the Manpower Accession Policy Working Group (MAPWG) and the Defense Advisory

<sup>2</sup> The Word Knowledge (WK) and Paragraph Comprehension (PC) test standard scores are combined to create a weighted Verbal (VE) composite. MK is the Math Knowledge test.

Table 2  
1990s Range-Corrected Validities for Navy ASVAB Composites With and Without Coding Speed Versus Final School Grade

Rating	N	Navy classification composite					(VE + MK + CS) – (VE + MK)
		VE + AR + MK	AR + 2MK + GS	VE + MK	VE + MK + CS		
Signalman	1,548	.56	.54	.56	<b>.59</b>	.03	
Radioman	2,263	<b>.62</b>	.61	.61	<b>.62</b>	.01	
Operations specialist	1,676	<b>.74</b>	.73	.72	<b>.74</b>	.02	
Dental technician	516	.63	.61	.62	<b>.64</b>	.02	
Personnel	942	.62	<b>.63</b>	.60	<b>.63</b>	.03	
Ship's service	801	<b>.49</b>	.48	.48	<b>.49</b>	.01	
Storekeeper	2,201	.65	.64	.63	<b>.66</b>	.03	
Aviation maintenance/administration	873	.72	.70	.71	<b>.74</b>	.03	
Aviation storekeeper	801	.63	.62	.63	<b>.65</b>	.02	
Mess management	2,589	<b>.65</b>	.63	.64	<b>.65</b>	.01	

*Note.* (1) The VE + MK composite of ASVAB tests was judged a suitable replacement for the services' administrative composites that contained CS (and in some cases, that also contained NO). (2) Validity coefficients were developed on final school grades that pertained to each Navy rating's initial technical training course. (3) Validity coefficients were corrected for range restriction using the PAY80 normative population correlation matrix as the unrestricted population from which, theoretically, future recruits would be selected using the ASVAB composites and cut scores (the ASVAB standards). ASVAB = Armed Services Vocational Aptitude Battery; VE = Verbal; AR = Arithmetic Reasoning; MK = Mathematics Knowledge; GS = General Science; CS = Coding Speed. Boldface indicated largest validity.

Committee-Military Personnel Testing (DAC-MPT) during the 1990s. The MAPWG consists of representatives from the services, the U.S. Military Enlistment Command, the DMDC, and the Office of the Secretary of Defense, Accession Policy. The MAPWG's responsibilities include resolving issues related to ASVAB test development, implementation and maintenance, and making policy recommendations. The DAC-MPT is an independent advisory group composed of volunteer experts in psychometrics, statistics, and test development. The DAC-MPT's responsibilities are to review the test development methods and calibration of the ASVAB and other military personnel selection and classification tests, but also to review validity results.

The composites of concern for the validity comparisons in Table 2 were the Navy's operational VE + MK + CS composite and the DoD suggested replacement, VE + MK. All composite tests were unit-weighted. The Navy composite with CS, showed, on average, .02 higher predictive validity than the VE + MK composite. A .02 increment in predictive validity may seem small, but in large-scale testing programs such as the ASVAB, can translate to substantial benefits both in terms of a reduction in training attrition and in associated costs (Schmidt, Dunn, & Hunter, 1995). For example, a .02 increment in predictive validity for training completion for the personnel selection scenario of 40% of ASVAB youth qualified for the occupation of air traffic controller would translate to a 1.5% expected improvement in the training completion rate given certain parameters (Taylor & Russell, 1939). These parameters could be, for example, (a) a 25% selection ratio (qualified youth resulting from the operational selection instrument with cut score), (b) an operational selection composite criterion-related (predictive) validity of .70 (predictive of final school grade in training that determines success and failure), (c) a .02 validity improvement for the candidate replacement composite, and (d) an observed 83% training completion rate (Taylor-Russell base rate .45 table). In this cost benefits scenario, at a \$100,000 training cost per enlistee and 1,000 recruited for the occupation, 15 fewer recruits would be expected to fail training merely due to the .02 validity increment in the selection composite. The expected cost savings for air traffic controllers under these conditions

would be \$1.5 million ( $\$100,000/\text{enlistee} \times 15$  enlistees). This amount of savings is for only one of many Navy ratings. Considerably higher cost-avoidance savings would occur if similar validity increments could be realized for several more ratings.

The Navy occupations (see Table 2) included a mix of clerical and nonclerical (e.g., signalman, radioman, operations specialist, and dental technician) ratings. These ratings are clearly different from mechanical ratings where AO is, on the face of it, more relevant (e.g., Aviation Mechanic). The variety of occupations in Table 2 is consistent with findings of the relevance of clerical speed tests in predicting performance for other than clerical occupations such as frogmen/SEAL (Alf & Gordon, 1957; Held, 2011) and air traffic controller (Held, 2006). An ASVAB classification composite that includes CS, VE + MK + MC + CS, is currently used by the SEALs and has been confirmed twice in Navy ASVAB validation/standards studies as the best predictor of success in the mentally challenging SEAL training (Held, 2011). In addition, an independent source outside the Navy confirmed the predictive validity of the VE + MK + MC + CS composite as optimal for the Navy SEALs for an entirely different dataset.<sup>3</sup>

The Navy's VE + MK + MC + CS composite used for SEAL classification is also used for the Navy's air traffic controller (AC) rating. Two predictive validation studies were conducted for the Navy air traffic controller rating that produced similar results despite the large difference in sample sizes ( $N_1 = 269$ ,  $N_2 = 71$ ; Held, 2006). In both air traffic controller studies, the VE + MK + MC + CS composite had the largest validity coefficient for predicting final school grades (with a tower operations hands-on performance measure component) and with about the same validity magnitude (in the .70 to .80 range). Also, in both cases, the CS composite showed about a .02 increment in validity over the highest validity ASVAB composites that did not contain CS. We note that not all ASVAB composites demonstrate validity magnitudes in the .70 to .80 range across military occupations (.25 to .85 for Navy) and that the benefits of a .02 validity increment depend on many factors including the baseline validity of the operational composite, the number of enlisted personnel required for the relevant oc-

cupations, the stringency of the cut score, and the observed failure rate.

The CS test also has demonstrated incremental validity for Army occupations. A study of many cognitive and noncognitive measures from the Army's Project A showed that the inclusion of CS among the predictors increased mean predicted performance across a broad set of occupations (Scholarios, Johnson, & Zeidner, 1994).

It should be noted that the predictive validity of CS may be moderated by job complexity. Schmidt et al. (1995) observed that perceptual measures such as CS do not provide predictive validity over a general ability factor of the ASVAB in low-complexity occupations but do for higher-complexity occupations. This finding has relevance for improving the military's classification systems by limiting the use of the CS test for assignment to only moderate to high-complexity occupations where the validity warrants. The question becomes how to use measures like CS in occupational classification when (a) the job is complex due to a requirement for technical knowledge in areas that are frequently updated and good reading comprehension skills in order to quickly understand technical manuals and (b) when several occupations are competing for recruits with high ASVAB scores.

It also should be noted that performance on the CS test under low-stakes conditions may be a function of motivation as well as ability. The AFQT is obviously a high-stakes military selection hurdle, as it determines enlistment eligibility, whereas the ASVAB classification composites are likely perceived as less high-stakes, as they do not affect enlistment qualification, only job assignment. Segal (2012) examined ASVAB and CS (then an ASVAB test) data from a nationally representative sample of 12,000 participants in the 1979 Longitudinal Survey of Youth study (for information on the NLSY go to <http://www.bls.gov/nls/nlsy79.htm>), where no high-stakes decisions were to be made. Participants were surveyed annually after testing regarding their earnings until 1994 and

<sup>3</sup> "Follow on Research Findings" submitted by Gallup Consulting, Inc., in 2011 to Director, Naval Special Warfare Recruiting Directorate, NAVSPECWARCEN, San Diego, CA.

biannually afterward. Results indicated that CS scores were significantly correlated with future earnings of study participants both by themselves and after controlling for cognitive ability (e.g., AFQT, educational attainment). Segal postulated that CS measures an underlying intrinsic motivational component related to test-taking performance and attainment of higher income levels over time. The identification and retention of individuals likely to remain motivated over time is of particular interest to the military.

**Assembling Objects.** As with CS, the AO test has shown small (about .02), but consistent, incremental validity when used in combination with other ASVAB tests. The .02 incremental validity results appear robust as they have been observed for several military occupations and performance criteria in studies conducted by the Army (Anderson et al., 2011; Russell, Le, & Putka, 2007), Marine Corps (Carey, 1994), and Navy (Held, Fedak, Crookenden, & Blanco, 2002; Held et al., 2004).

Table 3 shows the incremental validity of AO for predicting final school grades in various ASVAB composites during the timeframe that the Navy was evaluating both CS and AO for occupational classification (Held et al., 2002). As with CS, the validities in Table 3 were corrected for range restriction on the ASVAB using the multivariate method (Lawley, 1943) but were not corrected for criterion reliability. A bootstrap method was used to reduce the influence of outliers (reporting the median corrected validity from the bootstrap distribution).

As shown in Table 3, AO demonstrated on average about a .02 validity increment across the group of Navy ratings when compared to the non-AO ASVAB composite that was determined to have highest validity. For example, for the parachute rigger rating, the “best” composite without AO (AR + MK + EI + GS) had a validity of .656. Substituting AO for GS (AR + MK + EI + AO) produced a .022 increment in validity (to .678). For the builder rating, the best composite without AO (AR + MC + AS) had a validity of .628. Substituting AO for AS (AR + MC + AO) yielded a .015 increment in validity (to .643).

There are several points to be made in the Navy builder example. First, the AR + MC + AS composite is a Navy operational classification composite (mechanical) that has clear over-

Table 3  
*Range Corrected Validities for Navy ASVAB Composites With and Without Assembling Objects Versus Final School Grade*

Rating	Rating description	Best ASVAB validity	Best ASVAB + AO validity	Validity difference
Aviation boatsman's mate (equipment) (N = 244)	Services hydraulics and arresting gear maintenance	VE + AR + MK + MC (.626)	AR + GS + AS + AO (.639)	.013
Builder (N = 339)	Performs wood and concrete construction	AR + MC + AS (.628)	AR + MC + AO (.643)	.015
Construction mechanic (N = 260)	Services gasoline and diesel engines	AR + MC + AS (.573)	AR + GS + AS + AO (.586)	.013
Parachute rigger (N = 293)	Rigs parachutes, maintains survival equipment	AR + MK + EI + GS (.656)	AR + MK + EI + AO (.678)	.020
Quartermaster (N = 250)	Steers ship; logs compass readings, tides, bearings, etc.	VE + AR + MK + MC (.750)	VE + AR + MK + AO (.762)	.012
Signalman (N = 149)	Operates assorted visual and communications devices	AR + MK + EI + GS (.542)	AR + MK + EI + AO (.577)	.035

*Note.* Validities were corrected for range restriction; differences were computed as best ASVAB plus AO composite minus the best ASVAB composite. The author recognizes that, as with any statistic, there are confidence intervals in validity differences between composites not reported. ASVAB = Armed Services Vocational Aptitude Battery; AO = Assembling Objects; VE = Verbal; AR = Arithmetic Reasoning; MK = Mathematics Knowledge; MC = Mechanical Comprehension; GS = General Science; AS = Auto and Shop Information; EI = Electronics Information.



lap in constructs by using both MC and AS (see Table 1 for a description of these tests). Replacing AS with AO not only reduces that construct overlap but improves the predictive validity of the composite and reduces adverse impact for some groups. Second, the builder rating, which includes skilled carpenters, plasterers, roofers, and painters, is an occupation that is found in the other services, so incorporating the AO test in the other services' classification composites should yield benefits to their classification systems as it has for the Navy.

Although Table 3 provides only brief descriptions of the limited number and types of occupations included in the AO validity analyses, the major duties listed show the relevance of spatial ability, in particular of the type measured by the AO items (form construction from pieces and connection point locations for form pieces). Similar AO validity results were observed in a study of Navy aviation mechanics ratings (Held et al., 2004), where the criterion again was final school grade. This final grade, however, was considered more representative of the job tasks as it incorporated hands-on laboratory performance measures that were scored on a continuous scale.

As mentioned, similar incremental validities for composites using the AO test have been reported by the Army (Anderson et al., 2011; Russell et al., 2007) and Marine Corps (Carey, 1994). These studies included a variety of occupations (e.g., Army: infantryman, armor crewman, military police, light wheel vehicle mechanic, health care specialist, and motor transport operator; Marine Corps: automotive and helicopter mechanics) and several job performance criteria including measures of hands-on performance, job knowledge, and final course grades.

The AO test was also a part of the more recent Army Select21 project that had the main objective of helping to ensure the acquisition of soldiers with the knowledge, skills, and abilities needed to perform the types of tasks envisioned in a transformed Army involved with future combat systems. The Select21 project included a future-oriented job analysis to support the development of experimental selection and classification predictor measures and performance criteria (Knapp & Tremble, 2007) mapped to a different mix of knowledge, skills, and abilities resulting from projected changes in

the Army force structure and job requirements. In this regard, Russell et al. (2007) evaluated the predictive utility of the AFQT, a unit-weighted ASVAB Technical composite (AS, MC, and EI), and the AO test in the Select21 predictive validation study. The sample size varied by analysis and consisted of 414–739 first-term enlisted soldiers. After correction for range restriction on the AFQT and criterion unreliability, the validities for predicting general technical proficiency were AFQT (.52), Technical (.48), and AO (.38). When all three scores were used together, AO provided about a .03 increment in validity over the AFQT and Technical combined scores (.57 vs. .54). Noting that the AO test demonstrated incremental validity beyond the AFQT and Technical scores used together, Russell et al. stated that “Spatial [AO] could be a useful predictor beyond the ASVAB, not just beyond AFQT” (p. 68).

We recognize that additional research is needed to examine the effect of implementing multiple nonverbal reasoning tests in military classification systems that are evaluated in concert with the CS and AO tests, as well as the existing ASVAB tests.

### Gender/Minority Group Score Differences

In addition to demonstrating predictive validity and incremental validity, one of the criteria regarding the development and use of tests for personnel measurement and selection/classification is that they demonstrate the same relations to occupational criteria for majority and minority groups (lack of predictive bias) and that group mean differences are minimized (lack of adverse impact). One of the arguments for adding tests to the ASVAB that do not rely on learned content is to reduce mean score differences between majority and minority groups on service classification composites (Drasgow et al., 2006; Wise et al., 1992). Wise and colleagues examined the sensitivity and fairness of service ASVAB classification composites, specifically those containing the technical tests, for many Air Force, Army, Marine Corps, and Navy technical occupations. They observed that the ASVAB technical composites were generally equally fair when comparing regression slopes and the resulting increases in mean criterion scores associated with increases in predictor scores across gender and racial



groups. However, adverse impact was noted to some extent for the minority group. As a result, Wise and colleagues recommended the services consider adding valid tests to the ASVAB (or to their classification systems) that reduced or eliminated barriers to occupational assignments. In response, the Navy adopted the AO test, for which mean score differences between the majority group (males and Whites) and each of several minority groups (females, racial/ethnic minority groups) were smaller compared to differences observed on the ASVAB technical tests. This section does not address regression slope differences (bias or fairness) between groups, only adverse impact defined as group mean differences in ASVAB, AO, and CS tests.

Table 4 (from Held et al., 2002) provides a gender and race/ethnic group breakout of mean score differences for the ASVAB tests, AO, and CS calculated as effect sizes for Navy accessions<sup>4</sup> before CS was eliminated from the battery (and during the AO evaluation phase).

Table 4 also shows the mean differences between Whites and racial/ethnic minority (African American, Hispanic, Asian, and Native American) groups broken out by gender expressed as effect sizes. Effect sizes were calculated as the difference between the majority (White) and the specific minority group mean divided by the pooled group standard deviation (*SD*). Cohen (1988) characterizes standardized mean score differences of .2 as small, .5 as moderate, and .8 as large. For this study, an effect size equal to or greater than |.5| was considered a meaningful impact.

The same test effect size patterns were observed for males and females across the race/ethnic groups (White being the common comparison group), suggesting cultural differences. African Americans had the largest number of effect size differences across tests and gender followed by Hispanics, and Asians. No meaningful effect size differences were found for Native Americans for either males or females. Not considering Native Americans further, auto/shop (AS) had the largest effect size, favoring Whites (males and females), for the three majority and minority group comparisons (White vs. African Americans, Hispanics, and Asians). The effect size difference (favoring Whites) was largest for African Americans (1.13 for males and 1.09 for females). In contrast, AO, when compared to AS, had trivial

effect sizes with the exception of African Americans, where the effect size was .58 for both males and females. Both CS and AO had smaller effect sizes than any of the technical knowledge tests. The effect size for CS was trivial across all groups and gender with the exception of a small .21 effect size in the comparison of White and African American males (favoring Whites).

Figure 1 graphically shows the effect sizes for the ASVAB and CS tests for gender. It is a graphical representation of the Table 4 data collapsed across all groups (males,  $N = 35,831$ ; females,  $N = 8,246$ ).

As shown in Figure 1, CS was the only test to favor females, nearly reaching the |.5| effect size criterion. This outcome is consistent with previous research showing that females outperform males on clerical speed/accuracy tests (Majeres, 1988) and processing speed tasks involving digits and letters (Roivainen, 2011). We note that this female advantage has not been found to extend to reaction time tasks where males have been shown to outperform females (Roivainen, 2011).

The small mean score differences for men and women observed for the CS and AO tests compared to those for the technical knowledge tests (GS, AS, MC, or EI) enables more women to qualify for a broad range of jobs with no loss of predictive validity or classification effectiveness. Although the AO test seems appropriate for many mechanical occupations as a substitute for AS, the AO test as yet has not been fully evaluated across all types of Navy occupations, but will be in the near future. The Air Force is in the process of evaluating the AO tests in a broader array of occupations.

We note that a case is not being made to eliminate the ASVAB technical tests (GS, AS, MC, and EI) but that it is possible to provide alternative ASVAB standards with low adverse impact tests (as the Navy has done with both AO and CS) that meet or exceed the validity of the technically saturated ASVAB composites. The technical knowledge tests have high utility in military classification because they measure

<sup>4</sup> The Held et al. (2002) data are for Navy accessions, not applicants. It is likely that group effect sizes for accessions are smaller than those for applicants due to selection on the ASVAB.

Table 4

*Effect Size Analysis for Gender and Race/Ethnic Groups (FY99 Navy Accession Population)*

ASVAB	Male effect sizes Whites (N = 22,230)				Female effect sizes Whites (N = 4,454)			
	Af. Am. N = 6,117	Hisp. N = 4,049	Asian N = 1,777	Nat. Am. N = 1,523	Af. Am. N = 1,911	Hisp. N = 1005	Asian N = 383	Nat. Am. N = 410
GS	*0.93	*0.68	*0.78	0.03	*0.87	*0.68	*0.53	0.16
AR	*0.70	0.31	0.09	0.03	*0.62	0.29	-0.07	0.10
VE	*0.65	*0.59	*0.73	-0.01	*0.66	*0.57	0.45	0.12
MK	0.19	0.04	-0.42	0.05	0.11	0.02	-0.41	0.06
MC	*0.93	0.43	0.43	-0.01	*0.83	0.42	0.34	-0.03
AS	*1.13	*0.73	*1.04	-0.11	*1.09	*0.84	*0.94	0.01
EI	*0.76	*0.52	0.46	-0.01	*0.68	*0.61	0.39	0.14
AO	*0.58	0.18	-0.04	-0.05	*0.58	0.22	0.02	-0.03
CS	0.21	0.10	-0.08	0.06	0.17	0.18	-0.10	0.07

*Note.* \* Denotes an effect size greater than  $|\cdot 5|$  (half a standard deviation),  $\cdot 5$  being considered moderate. Effect size was calculated as the major group mean (White) minus the minor group mean, the difference divided by the pooled groups' standard deviation. ASVAB = Armed Services Vocational Aptitude Battery; Af. Am. = African American; Hisp. = Hispanic; Nat. Am. = Native American; GS = General Science; AR = Arithmetic Reasoning; VE = Verbal; MK = Mathematics Knowledge; MC = Mechanical Comprehension; AS = Auto and Shop Information; EI = Electronics Information; AO = Assembling Objects; CS = Coding Speed.

not only knowledge of the subject matter relevant to training and jobs, but potentially experience and interest that results in motivated engagement in technical endeavors, which involves/enhances the learning process.

### Improved Classification

Improved classification was considered with respect to two objectives: (a) increasing assignment flexibility and (b) improved performance.

**Increased assignment flexibility.** During the time that CS was being evaluated for elimination from the ASVAB, the Navy was con-

cerned not only about lower predictive validity from losing CS and increased adverse impact, but also about losing differential assignment capability (Johnson, & Zeidner, 1991). Scholarios et al. (1994) showed that CS provided differential assignment capability as well as increased mean predicted performance (Brogden, 1951). Without the use of CS, the Navy was concerned that assignment flexibility would be restricted, resulting in an increased number of applicants who could not be assigned to jobs. The Navy's evaluation of differential assignment capability, described more fully later in this paper, took the form of simulating recruit assignments to Navy ratings using and not using CS and AO in ASVAB classification composites. The objective was to see how many recruits would not be assigned across all ratings given their yearly goals (school seats) under varying ASVAB, CS, and AO classification scenarios. In all scenarios, the cut scores established for the composites that used CS or AO and those that did not were effectively set to be the same for each Navy rating (recognizing that to achieve a better fill rate across ratings all one would need to do is to lower the cut scores, but with the expectation of lower performance).

Two classification algorithms were used in the Navy's recruit assignments to ratings simulation studies. The first was developed by Navy Personnel Research, Studies, and Technology (Folchi, 2007; Folchi & Watson, 1997) and

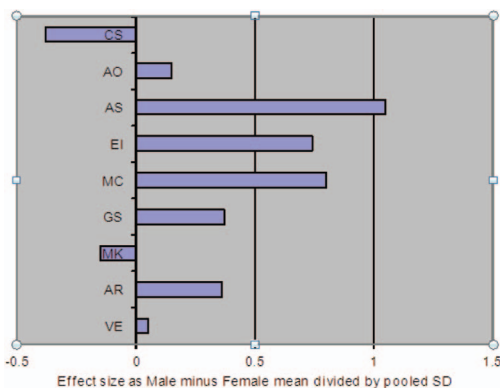


Figure 1. Fiscal Year 1999 Navy Armed Services Vocational Aptitude Battery effect sizes for gender. See the online article for the color version of this figure.

operationalized by EDS Federal Engineering and Logistics under contract (EDS Federal, 2001). The algorithm is now incorporated in the Navy's operational rating classification system called the Rating Identification Engine (Crookenden & Blanco, 2002). The algorithm's purpose is to generate a ranking of ratings (occupations or alternatively specific jobs) to which a person should be classified considering input personnel data and two utility functions. One utility function rewards applicants with high ASVAB composite scores relevant to a particular rating. The other function discourages classifying largely overqualified applicants to ratings considered not optimally challenging. Data for recruits are entered into the classification system in a sequential manner (which involves a random selection component) in order to mimic the operational assignment process (in contrast to assigning all recruits in a batch).

Four composite sets were created for the simulation. Composite Set 1 (baseline) contained no ASVAB composites with either CS or AO. Composite Set 2 contained some composites with AO where predictive validity warranted the test's use. Likewise, Composite Set 3 contained some composites with CS where the validity warranted the test's use. Composite Set 4 contained all of the CS and AO composites for their respective ratings along with the remaining ASVAB composites that applied to ratings where the CS and AO tests did not add incre-

mental validity. Only one composite per rating applied in all simulation scenarios and cut scores were set (on a rating's counterpart composite across scenarios) to qualify the same percentage of recruits. Differential assignment capability was defined as (a) the increase in the percentage of the recruit population "assigned" to ratings and (b) the lowest standard deviation of fill rate (indicating even fill).

Over 80 Navy ratings with their associated recruitment goals for 4 or 5 year enlistment training programs were involved in the study. All were referred to in the simulation as "jobs." Males and females were simulated in separate analyses because some jobs were not open to females. Finally, four scenarios were applied for the four sets of composites that varied the ratio of "job slots" to recruits to mimic different recruiting environments (e.g., either too many or not enough recruits for slots). Table 5 shows the results of the simulations.

Table 5 shows that, in each of the four classification simulation scenarios, providing an ASVAB composite set that included some composites with the AO or CS tests, and especially with both, resulted in fewer recruits "unassigned" to jobs. Also, the standard deviation of job fill (indicating evenness of distribution) tended to decrease with the addition of composites with the AO and CS tests. The obvious exception is for scenario 4 (13.4% more male jobs than males to assign) where *everyone* was

Table 5  
Rating Classification Simulation Results

	Composite set without AO or CS	Composite set with AO	Composite set with CS	Composite set with AO and CS
Scenario 1: 1.7% less female jobs than females (8,134 jobs; 8,275 females)				
Unassigned recruits	469	413	389	288–303 (range with 4 runs)
Job fill <i>SD</i>	16.1%	15.1%	14.6%	13.7–14.8 %
Scenario 2: 2.5% more female jobs than females (8,484 jobs; 8,275 females)				
Unassigned recruits	501	440	279	279–300 (range with 4 runs)
Job fill <i>SD</i>	20.2%	18.7%	16.7%	16.4–17.4 %
Scenario 3: 6.2% more male jobs than males (38,402 jobs; 36,154 males)				
Unassigned recruits	938	661	785	492–555 (range with 4 runs)
Job fill <i>SD</i>	13.8%	13.4%	13.0%	12.6–14.4 %
Scenario 4: 13.4% more male jobs than males (40,995 jobs; 36,154 males)				
Unassigned recruits	387	71	213	0 (range with 4 runs)
Job fill <i>SD</i>	15.9%	15.6%	15.6%	18.2–19.3 %

Note. AO = Assembling Objects; CS = Coding Speed; *SD* = standard deviation.

assigned to a job using the AO and CS composite set, but with less of an even fill across ratings (standard deviations ranged from 18.2% to 19.4% for 4 runs, as compared to the high of about 15% for the other runs).

The other Navy sequential assignment simulation application, developed by the Lewin Group, Inc. (Hogan & Simonson, 2004) is used as a decision tool when conducting Navy ASVAB validation/standards studies. The algorithm assigns a recruit (drawn randomly from a prior year's recruit population that is an input file to the application) to a rating (with annual goals or slot numbers for each rating provided in a separate input file) for which the difference between that recruit's ASVAB composite score is a minimum when compared to the rating's operational composite's cut score. Despite this minimum ASVAB delta criterion for a rating assignment (which involves a random tie breaker routine when ties occur), all ratings end up with a fairly wide range (distribution) of ASVAB scores to the right of the cut score because there are a limited number of recruits with ASVAB composite scores at the margin.

Consistent with the operational Rating Identification Engine algorithm simulation study, the Lewin Group, Inc. (Hogan & Simonson, 2004) study also showed more recruits in the aggregate assigned to ratings when the CS and AO composites were used compared to when they were not.<sup>5</sup> Not only was differential assignment improved (i.e., more recruits being assigned to ratings at the same relative cut scores), but at lower recruiting costs. The lower recruiting costs were due to a lower proportion of high AFQT recruits with a high school diploma required to fill the ratings due to the lower correlation of CS and AO to the AFQT tests used operationally for military selection. Traditionally, the higher AFQT youth with a high school diploma are more expensive to recruit.

**Improved performance.** Aside from improved flexibility, another person–job-match goal considered was improved performance. Horst (1954, 1955, 1956) and Brogden (1946, 1951, 1955, 1959) were early pioneers in recognizing that an important outcome of improved classification, termed CE, could be measured in terms of the overall performance of those classified into jobs. The Army has been the main driver among the services in research

showing improvements in CE to optimize person–job match under the framework of DAT that considers many aspects of classification effectiveness (Johnson & Zeidner, 1995). In a study involving many cognitive and noncognitive measures from the Army's Project A, Scholarios et al. (1994) showed that the inclusion of CS among the predictors increased mean predicted performance (MPP) across a broad set of occupations and improved differential assignment capability. In one experimental test battery that involved the largest number of occupations, CS, in an optimally derived equation, was selected first based upon its differential assignment index. Although the Navy did not use the same statistical methods for assessing CS differential assignment capability as did Scholarios et al. (1994), their recruit occupational assignment simulations demonstrated that lowering the average intercorrelation of the ASVAB tests (which inclusion of CS does) will improve the breadth of coverage of the cognitive domain.

Even from an economist's viewpoint, there is recognized utility in increasing the differential classification capability of the performance predictor variables when considering assignment algorithms. For example, Schmitz and Holz (1987) were keenly aware of the differences in selection and classification in their personnel assignment models:

Selection focuses on the differences among individuals, generally using a single scale of value or utility. Applicants are classified into two categories: those satisfactory for employment and those not . . .

Differential classification deals with differences within an individual with respect to various skills. A particular individual may have a high aptitude for mathematics but poor writing skills. Another may have considerable talent for electronics jobs but poor communication ability. Classification requires the use of two or more different performance predictors. (p. 440)

Johnson and Zeidner (1991, 1995) developed the principles of DAT that considered the tenets of Brogden's (1959) measure of CE but also the formation of more or less homogeneous occu-

<sup>5</sup> The Lewin Group, Inc., application is used for many purposes, including comparing diversity across ratings to the operational classification state and also to assess the fill potential for women in technical ratings (females historically score lower on the ASVAB technical tests).



pational groups upon which to develop prediction equations that maximize MPP in the aggregate. CE equates to MPP and captures both predictive validity and the intercorrelation of the ordinary least squares (OLS) equation estimates of performance (job or training). The CE formulas that are outgrowths of the early Army pioneers' work are presented succinctly by Statman, Gribben, Naughton, and McCloy (1998, p. 7), as follows, where MPP equates to CE:

$$\text{MPP} = R(1 - r)^{1/2} Z_m \quad (1)$$

where

MPP = mean predicted performance standard score of a group of applicants assigned to  $m$  jobs,

$R$  = average predictive validity of OLS estimates for all jobs,

$r$  = average intercorrelation of the OLS estimates, and

$Z_m$  = mean criterion standard score of the group after assignment to the  $m$  jobs with equal vacancies (called quotas).

Statman et al. (1998, p. 7) noted that  $R$  (predictive validity) is positively related to CE. However, to maximize the CE index of MPP, the intercorrelations within the prediction equations across occupations need to be lowered. Both CS and AO help to lower the average ASVAB test intercorrelations and so also the intercorrelations of OLS prediction equations that contain them. Further applied research using the principles developed by the classification theorists could be applied to the ASVAB that includes AO, CS, and all existing and new measures potentially appropriate for the ASVAB or as adjuncts. This work could be done considering all of the performance predictors in concert, not in the piecemeal fashion that the Navy first took in their evaluation of the benefits of the AO and CS tests.

## Discussion

Drasgow et al. (2006) recommended that cognitive content be added to the ASVAB that was not dependent on verbal skills or acquired knowledge. To this end, we examined both theoretical and empirical support for two tests, CS (speed/accuracy) and AO (spatial) for use in military occupational classification. We focused on three factors of particular importance to the

U.S. military: (a) incremental predictive validity when these tests are used in combination with the ASVAB academic and technical knowledge tests for predicting important occupational criteria, (b) reducing subgroup differences (adverse impact) for women and some minority groups, and (c) improved classification, both in terms of increased assignment flexibility and improved performance, due to enhanced differential assignment capability.

Our comprehensive evaluation approach was consistent with the recommendation from Drasgow et al. (2006, p. 25) that evaluations of the benefit of adding a test to the ASVAB go beyond examination of incremental validity to include reduced adverse impact and improved classification efficiency. Our analyses indicated that, in aggregate, inclusion of AO and CS in Navy classification composites provided small increments in predictive validity, reduced adverse impact for women and some minority groups, and improved classification in terms of increased assignment flexibility and, theoretically according to formal CE methods, improved mean performance.

## Incremental Validity

Despite the AO and CS tests being based on different concepts of cognitive ability and content domains than the ASVAB verbal, math, and technical knowledge tests, only small increments in predictive validity (about .02 on average) were observed in the studies reported in this paper. The overall mean incremental predictive validity for CS and AO was close to that estimated by Schmidt et al. (1995), who acknowledged that although the .02 validity increment seems small, it has the potential to produce substantial cost savings in large scale testing programs such as the ASVAB (used for military occupational classification).

Given the theoretical and empirical evidence supporting the use of the AO and CS tests, it is instructive to discuss the reason for the observed small .02 increment in validity these tests contribute to the ASVAB verbal, math, and technical knowledge tests. To some extent, the small incremental predictive validity for CS and AO is simply that all cognitive tests measure to varying extents general mental ability. Confirmatory factor analyses have shown that both CS (Ree & Carretta, 1994) and AO (Drasgow et al.,



2006) contribute strongly to a general factor derived from all of the ASVAB tests and interpreted as psychometric *g*. These results are consistent with studies that have compared cognitive tests developed from different theoretical approaches only to find they mainly measured *g* (e.g., Keith, Kranzler, & Flanagan, 2001; Stauffer, Ree, & Carretta, 1996). One explanation for the high degree of overlap in what is being measured by tests developed on different theoretical bases (such as Gc and Gf) and the observed low incremental validity for CS and AO in the studies reviewed in this paper can be found in what Spearman (1923) called the “indifference of the indicator” or “indifference of the fundamentals.” This means that test *content* is not fundamental to the measurement of *g*. The test content is the vehicle that allows for the expression of relationships and differences measured in cognitive ability tests. Cognitive ability can and has been measured with verbal, quantitative, and spatial test items of widely varying content. Measures of nonverbal reasoning (e.g., Raven’s Matrices; Raven, 1939) have no verbal or quantitative content, and little spatial content yet measures *g*. Chronometric measures (i.e., cognitive speed) that require no verbal, quantitative, or spatial content also have been shown to measure *g* (Jensen, 2006).

Despite the viewpoint of a predominant *g* factor, there is utility to tests like CS and AO in the large-scale ASVAB testing program because, despite these tests’ low incremental validity, they apply to many occupations and over time many individuals will be better fit to occupations with the expectation of better training and by extension, job performance. Also, use of the CS and AO tests unquestionably lowers adverse impact or test score barriers.

### Adverse Impact

When evaluating new predictors, the primary focus of much of the research has been on their incremental validity over existing predictors (e.g., Besetsny, Earles, & Ree, 1993; Wolfe, 1997; Wolfe et al., 1997). While incremental validity is desirable when adding test content to a battery such as the ASVAB, it may not be necessary if the new measure can replace an existing test with no loss of validity but provide some other benefit such as demonstrating smaller mean subgroup differences. As was

seen, smaller mean subgroup differences can expand the occupation-qualified recruit pool.

The most significant contributions of the CS and AO tests to assignment of military enlisted personnel to occupations may be their potential to reduce adverse impact for females and some minority groups. The potential to reduce adverse impact is due to less reliance on learned content than is the case for the ASVAB verbal, math, and technical knowledge tests. The Army recognized the CS test’s benefit in this regard during an exploration of methods in measuring ASVAB composite fairness after the NO and CS tests were eliminated. Based on their evaluation, Zeidner, Johnson, Vladimirovsky, and Weldon (2004) recommended that the CS test be restored to the ASVAB to reduce adverse impact and increase MPP.

### Improved Classification Outcomes

The potential of CS and AO to improve classification flexibility by increasing the number of occupations recruits can qualify for without compromising aggregate expected performance is part of improving CE and was shown in this paper through two Navy rating assignment simulation studies. Limited work thus far has been done by the Navy to address the formal tenants and formulas the Army uses to assess CE within the DAT framework. Applied work will be planned to do so when test score data are available for all of the tests scheduled for evaluation as formal ASVAB tests or as ASVAB adjuncts and as military performance data become available for comprehensive test validation studies. However, the work done by the Army at least supports the expectation that the inclusion of CS would enhance the outcome of mean predicted performance.

Given the various Army’s and Navy’s independent assessments of the CS and AO tests and the recent DMDC finding of a more robust CS test than previously thought, the next step could be the services participating in a joint-service evaluation of the two tests (on the three criteria presented in this paper). Taken together, the empirical support for the CS and AO tests indicates multiple benefits to the services in (a) increasing their ability to predict which recruits will perform well in training and thus establish more effective cognitive ability cut scores, (b) lowering score barriers for women and some

minority groups in qualifying for important occupations and thus improve the diversity that is highly valued in the military, and (c) increasing the overall percentage of recruits qualified for classification/assignment to occupations with the same previously expected success rates. All of the services currently are administered the AO test in the ASVAB Enlisted Testing Program (both CAT-ASVAB and paper-and-pencil ASVAB) and it should not be much of an additional testing burden to administer the 7-min CS test at the MEPS to all service applicants. The only additional burden would be to ensure high quality levels for the criterion used for the validity/incremental validity and CE portions of the analysis.

### New Test-Item Development

Another advantage of CS and AO that has not yet been discussed and that applies to other measures that do not use academic or knowledge-based content is the relative ease and lower expense of new test-item development (compared with measures of learned content). That is, tests like AO and CS and the candidate ASVAB measures of nonverbal reasoning and working memory can be developed as automated item-generated tests. Automated item-generated tests have the potential to (a) reduce item writer requirements and test maintenance levels compared to that required for academic and knowledge based tests, (b) alleviate the need to continually review test content for outdated material, and (c) minimize the monitoring efforts needed to identify test security and test compromise issues. Further, tests like AO, CS, and measures of nonverbal reasoning and working memory can be made adaptive like the ASVAB tests, therefore reducing testing time. Conversion to an adaptive format could also improve measurement precision but requires a psychometric balancing act.

### Limitations and Future Studies

There are limitations to the analyses reported in this paper supporting the CS and AO tests that include (a) some occupations having small sample sizes, (b) not having a large variety of very different types of occupations, (c) not evaluating the inclusion of both AO and CS in the same classification composites, and (d) not hav-

ing all potential or adjunct ASVAB tests to compare outcomes.

As noted earlier, DMDC and the services plan to evaluate new measures of nonverbal reasoning and working memory for potential inclusion on the ASVAB or as special tests (adjuncts to the ASVAB). These tests include Table Processing (Segall, 2010), a purer measure of processing speed than CS and a possible replacement or addition, and an updated version of the ECAT Mental Counters (MCt) test, which measures working memory. The ECAT MCt test demonstrated incremental validity for predicting military air traffic controller performance (Held & Wolfe, 1997) and is expected to contribute to an operational Navy ASVAB composite in 2015. MCt also is expected to be useful for other military occupations for which working memory has been identified as important (e.g., unmanned aerial vehicle operators; Paullin, Ingerick, Trippe, & Wasko, 2011). Further, MCt has been included in the new version of the Defense Language Aptitude Battery, which will be further evaluated for predicting foreign language learning outcomes in a validity confirmation study.

It may be that recent changes in technology and job requirements in the military (e.g., wider application of unmanned aerial vehicle platforms and cyber/IT) may show even more relevance for CS, AO, and the other tests now being evaluated for military occupational classification. The new performance predictor tests at the very least are expected to have benefits similar to those demonstrated for CS and AO in this paper. That is, the nonverbal reasoning and working memory tests are expected to show at least small amounts of incremental validity to current ASVAB tests, reduce adverse impact for females and some minority groups in qualifying for some occupations, and increase the proportion of recruit populations qualified for occupations on a cognitive basis.

### References

- Abrahams, N., & Alf, E. F. (2001). *The equivalency of operational ASVAB and experimental measures of perceptual speed across alternative modes of administration*. (Personnel Decisions Research Institutes, Inc. Contract GS10F0059K, Order No. N0063900FA156). Millington, TN: Navy Personnel Research Studies and Technology.

- Abrahams, N., Walton-Paxton, E., Alf, E. F., Barton, M. A., Cole, D. R., & Kieckhafer, W. G. (1996). *Hardware independent measures of the ASVAB clerical proficiencies test*. (RGI, Inc. Contract N66001-95-D-960, Order No. 7J06). San Diego, CA: Navy Personnel Research and Development Center.
- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6, 259–290. doi:10.1037/1076-898X.6.4.259
- Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology*, 9, 5–37. doi:10.1207/s15327876mp0901\_1
- Alf, E. P., & Gordon, L. V. (1957). *A fleet validation of selection tests for underwater demolition team training*. (FRASD Memo 57–3). San Diego, CA: U.S. Naval Personnel Research Field Activity.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, L. M., Hoffman, R., Tate, B., Jenkins, J., Parish, C., Stachowski, A., & Dressel, J. D. (2011). *Assessment of assembling objects (AO) for improving predictive performance of the Armed Forces Qualifications Test*. (Tech. Rep. No. 1282). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Besetsny, L. K., Earles, J. A., & Ree, M. J. (1993). Little incremental validity for a special test of abstract reasoning in an Air Force intelligence career field. *Educational and Psychological Measurement*, 53, 993–997. doi:10.1177/0013164493053004012
- Bloxom, B., Thomasson, G., Wise, L., & Branch, R. (1993). *Initial operational test and evaluation of new optical-mark-reader answer sheets for the Armed services Vocational Aptitude Battery (ASVAB)*. (ADA 274 190). Monterey, CA: Defense Manpower Data Center.
- Brogden, H. E. (1946). An approach to the problem of differential predictions. *Psychometrika*, 11, 139–154. doi:10.1007/BF02289295
- Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement*, 11, 173–195. doi:10.1177/001316445101100201
- Brogden, H. E. (1955). Least squares estimates and optimal classification. *Psychometrika*, 20, 249–252. doi:10.1007/BF02289020
- Brogden, H. E. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement*, 19, 181–190. doi:10.1177/001316445901900204
- Buscigilo, H. H., & Palmer, D. R. (1996). *An empirical assessment of coaching and practice effects on three Army tests of spatial aptitude*. (ARI Research Note 96–70). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Buscigilo, H. H., Palmer, D. R., King, I. H., & Walker, C. B. (1994). *Creation of new items and forms for the Project A Assembling Objects test*. (Tech. Rep. No. 1004). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Campbell, J. P., & Zook, L. M. (Eds.). (1992). *Building and retaining the career force: New procedures for screening and assigning Army enlisted personnel: Annual report, 1990 fiscal year*. (ARI Tech. Rep. No. 952). Alexandria, VA: U.S. Research Institute for the Behavioral and Social Sciences.
- Carey, N. B. (1994). Computer predictors of mechanical job performance: Marine Corps findings. *Military Psychology*, 6, 1–30. doi:10.1207/s15327876mp0601\_1
- Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In B. H. Kantowitz (Series Ed.), P. S. Tsang, & M. A. Vidulich (Vol. Eds.), *Human factors in transportation: Principles and practices of aviation psychology* (pp. 357–396). Mahwah, NJ: Erlbaum.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale, NJ: Erlbaum.
- Crookenden, M., & Blanco, T. (2002). *SCORE test, comparing composite tests*. (Delivery Order EDS 02–002). EDS draft technical report prepared for Navy Personnel Research, Studies, and Technology, Millington, TN.
- Dockeray, F. C., & Isaacs, S. (1921). Psychological research in aviation in Italy, France, England, and the American Expeditionary Forces. *Journal of Comparative Psychology*, 1, 115–148. doi:10.1037/h0070608
- Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed services Vocational Aptitude Battery (ASVAB), FR-06–25*. Alexandria, VA: Human Resources Research Organization.
- Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualifying Test form S: Analysis and comparison

- with previous forms. *Military Psychology*, 22, 68–85. doi:10.1080/08995600903249255
- EDS Federal. (2001). *Alternative algorithms for job selection*. (Contract Number: N66001-00-F-CR00/GS-35F-0015K, Delivery Order: FISC NOR [0014]); Sabre (GSA 01-002). Millington, TN: Navy Personnel Research, Studies, and Technology.
- Fleishman, E. A., Quaintance, M. K., & Broadling, K. A. (1994). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.
- Folchi, J. (2007). *RIDE vs. CLASP comparison and evaluation: Models and parameters*. (NPRST-TR-07). Millington, TN: Navy Personnel Research, Studies, and Technology.
- Folchi, J., & Watson, S. (1997). *Specifications for RIDE utility functions*. Unpublished technical report. Millington, TN: Navy Personnel Research, Studies, and Technology.
- Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York, NY: Wiley.
- Held, J. (2006). *Armed services Vocational Aptitude Battery (ASVAB) standards: Air traffic control rating*. (NPRST Letter Report Ser 3900, PERS-1/00047). Millington, TN: Navy Personnel, Research, Studies, and Technology.
- Held, J. (2011). *Armed services Vocational Aptitude Battery (ASVAB) standards: Special warfare operations (SO) SEAL from program to rating*. (NPRST Letter Report Ser 3900, BUPERS-/0009211). Millington, TN: Navy Personnel, Research, Studies, and Technology.
- Held, J. D., Fedak, G. E., Crookenden, M. P., & Blanco, T. A. (2002). *Test evaluation for augmenting the Armed services Vocational Aptitude Battery*. Proceedings of the 44th International Military Testing Association (pp. 291–297). Ottawa, Canada.
- Held, J., Fedak, G., & Johns, C. (2004). *Armed services Vocational Aptitude Battery (ASVAB) standards: Aviation mechanics*. (NPRST Letter Report Ser 3900). Millington, TN: Navy Personnel Research, Studies, and Technology.
- Held, J. D., & Foley, P. P. (1994). Explanations for accuracy of the general multivariate formulas for correcting for range restriction. *Applied Psychological Measurement*, 18, 355–367. doi:10.1177/014662169401800406
- Held, J. D., & Wolfe, J. H. (1997). Validities of unit-weighted composites of the Armed services Vocational Aptitude Battery (ASVAB) and Enhanced Computer Administered Tests (ECAT). *Military Psychology*, 9, 77–84. doi:10.1207/s15327876mp0901\_4
- Helme, W. H., Graham, W. R., & Anderson, A. A. (1962). *Development of ACB automotive information and clerical speed tests, forms 3 and 4*. (USAPRO Tech. Rep. No. 1122). Washington, DC: U. S. Army Personnel Research Office.
- Hogan, P., & Simonson, B. (2004). *Selection and classification cost effectiveness model*. (Contract Number: MOBIS N00639-02-F-A-097). The Lewin Group Inc. software and user manual prepared for Navy Personnel Research, Studies, and Technology, Millington, TN.
- Holzinger, K. J., & Harman, H. H. (1941). *Factor analysis*. Chicago, IL: University of Chicago Press.
- Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychological Monographs: General and Applied*, 68, 1–31. doi:10.1037/h0093676
- Horst, P. (1955). A technique for the development of a multiple absolute prediction battery. *Psychological Monographs: General and Applied*, 69, 1–22. doi:10.1037/h0093687
- Horst, P. (1956). Multiple classification by the method of least squares. *Journal of Clinical Psychology*, 12, 3–16. doi:10.1002/1097-4679(195601)12:1<3::AID-JCLP2270120102>3.0.CO;2-U
- Hull, C. L. (1928). *Aptitude testing*. Yonkers-on-Hudson, NY: World Book. doi:10.1037/11019-000
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, UK: Elsevier.
- Johnson, C. D., & Zeidner, J. (1991). *The economic benefits of predicting job performance: Vol. 2: Classification efficiency*. New York, NY: Praeger.
- Johnson, C. D., & Zeidner, J. (1995). *Differential assignment theory sourcebook*. (ARI Research Note 95–43). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3rd ed.). *School of Psychology Review*, 30, 89–119.
- Knapp, D. J., & Tremble, T. R. (Eds.). (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (ARI Tech. Rep. No. 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lawley, D. (1943). A note on Karl Pearson's selection formula. *Royal Society of Edinburgh, Proceedings, Section A*, 62, 28–30.
- Majeres, R. L. (1988). Serial comparison processes and sex differences in clerical speed. *Intelligence*, 12, 149–165. doi:10.1016/0160-2896(88)90013-X
- McWhite, P., & Greenston, P. M. (1998). *Design considerations for the Enlisted Personnel Allocation*



- tion System (EPAS) in its interface with the Army Recruit Quota System (REQUEST) (Study Note 98-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Paullin, C., Ingerick, M., Trippe, D. M., & Wasko, L. (2011). *Identifying best bet entry-level selection measures for U.S. Air Force remotely piloted aircraft (RPA) pilot and sensor operator (SO) occupations* (AFCAPS-FR-2011-0013). Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Peterson, N. G., Russell, T. L., Hallam, G., Hough, L., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery. In J. P. Campbell & L. M. Zook (Eds.), *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel: Annual Report 1990 FY* (pp. 73-199). (Research Note 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Raven, J. C. (1939). *Progressive matrices: A perceptual test of intelligence*. London, UK: H. K. Lewis.
- Ree, M. J., & Carretta, T. R. (1994). Factor analysis of ASVAB: Confirming a Vernon-like model. *Educational and Psychological Measurement*, 54, 459-463. doi:10.1177/0013164494054002020
- Ree, M. J., & Wegner, T. G. (1990). Correcting differences in answer sheets for the 1980 Armed services Vocational Aptitude Battery reference population. *Military Psychology*, 2, 157-169. doi:10.1207/s15327876mp0203\_3
- Roivainen, E. (2011). Gender differences in processing speed: A review of recent research. *Learning and Individual Differences*, 21, 145-149. doi:10.1016/j.lindif.2010.11.021
- Rumsey, M. G. (2012). Military selection and classification in the United States. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford handbook of military psychology* (pp. 129-147). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195399325.013.0054
- Russell, T. L., Le, H., & Putka, D. (2007). Predictor measure validation methods and Armed services Vocational Aptitude Battery results. In D. J. Knapp & T. R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures*, ARI Tech. Rep. No. 1205. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Russell, T. L., & Peterson, N. G. (2001). The experimental battery: Basic attribute scores for predicting performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 269-306). NY: Erlbaum.
- Russell, T. L., Peterson, N. G., Rosse, R. L., Hatten, J. T., McHenry, J. J., & Houston, J. S. (2001). The measurement of cognitive, perceptual, and psychomotor abilities. In J. P. Campbell & D. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 71-109). Hillsdale, NJ: Erlbaum.
- Sager, C. E., Peterson, N. G., Oppler, S. H., Rosse, R. L., & Walker, C. B. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology*, 9, 97-120. doi:10.1207/s15327876mp0901\_6
- Schmidt, F. L., Dunn, W. L., & Hunter, J. E. (1995). *Potential utility increases from adding new tests to the Armed services Vocational Aptitude Battery (ASVAB)* (NPRDC-TN-95-5). San Diego, CA: Navy Personnel Research and Development Center.
- Schmitz, E. Z., & Holz, B. W. (1987). Technologies for person-job matching. In J. Zeidner (Ed.), *Human productivity enhancement* (Vol. 2, pp. 439-472). New York, NY: Praeger.
- Scholarios, D., Johnson, C., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology*, 79, 412-424. doi:10.1037/0021-9010.79.3.412
- Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science*, 58, 1438-1457. doi:10.1287/mnsc.1110.1509
- Segall, D. O. (1997). The psychometric comparability of computer hardware. In W. A. Sands, B. K. Waters, & J. R. McBride, (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 219-226). Washington, DC: American Psychological Association.
- Segall, D. O. (2004). *Development and evaluation of the 1997 ASVAB score scale* (Tech. Rep. No. 2004-002). Seaside, CA: Defense Manpower Data Center.
- Segall, D. O. (2010). *Table processing speed: A new processing speed test*. Presentation given at the Oct. 19, 2010 Technical Committee of the Manpower Accession Policy Working Group. Seaside, CA: Defense Manpower Data Center.
- Segall, D. O., Moreno, K. E., Bloxom, B. M., & Hetter, R. D. (1997). The psychometric procedures for administering CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride, (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 219-226). Washington, DC: American Psychological Association.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: SIOP.



- Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London, UK: Macmillan.
- Statman, M. A. L., Gribben, M. A., Naughton, J. A., & McCloy, R. A. (1998). *The development of a new research paradigm for studying aptitude-treatment interactions* (HumRRO FR-WATSD-97-25). Alexandria, VA: Human Resources Research Organization.
- Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen's analyses. *Journal of General Psychology*, 123, 193-205. doi:10.1080/00221309.1996.9921272
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23, 565-578. doi:10.1037/h0057079
- Thurstone, L. L. (1938). *Primary mental abilities*. *Psychometric Monographs No. 1*. Chicago, IL: University of Chicago Press.
- Thurstone, L. L., & Thurstone, T. G. (1941). *The Primary Mental Abilities Tests*. Chicago, IL: Science Research Associates.
- Weltin, M. M., & Popelka, B. A. (1983). *Evaluation of the ASVAB 8/9/10: Clerical composite for predicting training school performance* (Tech. Rep. No. 594). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D. (1992). *Sensitivity and fairness of the Armed services Vocational Aptitude Battery (ASVAB) technical composites* (DMDC Tech. Rep. No. 92-002). Monterey, CA: Defense Manpower Data Center.
- Wolfe, J. H. (1997). Incremental validity of ECAT battery factors. *Military Psychology*, 9, 49-76. doi:10.1207/s15327876mp0901\_3
- Wolfe, J. H., Alderton, D. L., Larson, G. E., Bloxom, B. M., & Wise, L. L. (1997). Expanding the content of CAT-ASVAB: New tests and their validity. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 239-249). Washington, DC: American Psychological Association.
- Zeidner, J., & Johnson, C. D. (1991). *The economic benefits of predicting job performance, Vol. 3: Estimating the gains of alternative policies*. New York, NY: Praeger.
- Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 377-410). Hillsdale, NJ: Erlbaum.
- Zeidner, J., Johnson, D., Vladimirovsky, Y., & Welton, S. (2004). *Comparison of alternative methods of measuring ASVAB test composite fairness* (Study Note 2004-06). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

### E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!